1.225J (ESD 225) Transportation Flow Systems

Lecture 8

Delays in Probabilistic Models: Elements from Queueing Theory

Profs. Ismail Chabini and Amedeo Odoni

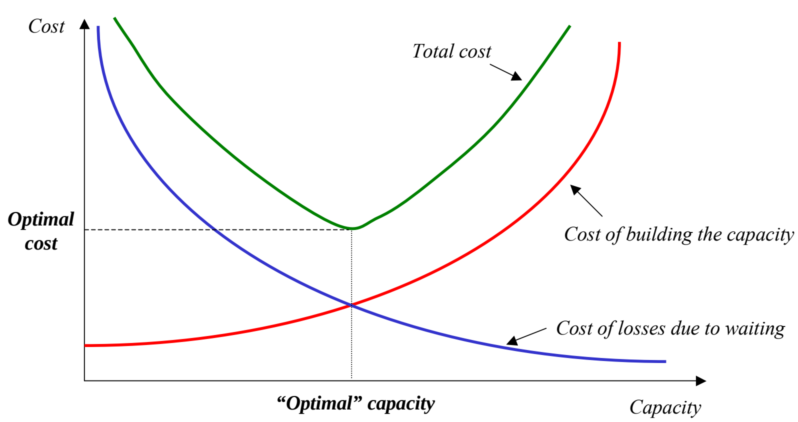
Lecture 8 Outline

- Introduction to Queueing
- Conceptual Representation of Queueing Systems
- Codes for Queueing Models
- Terminology and Notation
- Little's Law and Basic Relationships
- Exponential Distribution for Interarrival and Service times Modeling
- State Transition Diagram
- Derivation of waiting characteristics for *M/M/1*
- Summary

Applications of Queueing Theory

- Some familiar queues:
  - Airport check-in
  - Automated Teller Machines (ATMs)
  - Fast food restaurants
  - On hold on an 800 phone line
  - Urban intersection
  - Toll booths
  - Aircraft in a holding pattern
  - Calls to the police or to utility companies
- Level-of-service (LOS) standards
- Economic analyses involving trade-offs among operating costs, capital investments and LOS
- Queueing theory predicts various characteristics of waiting lines (or queues) such as average waiting time

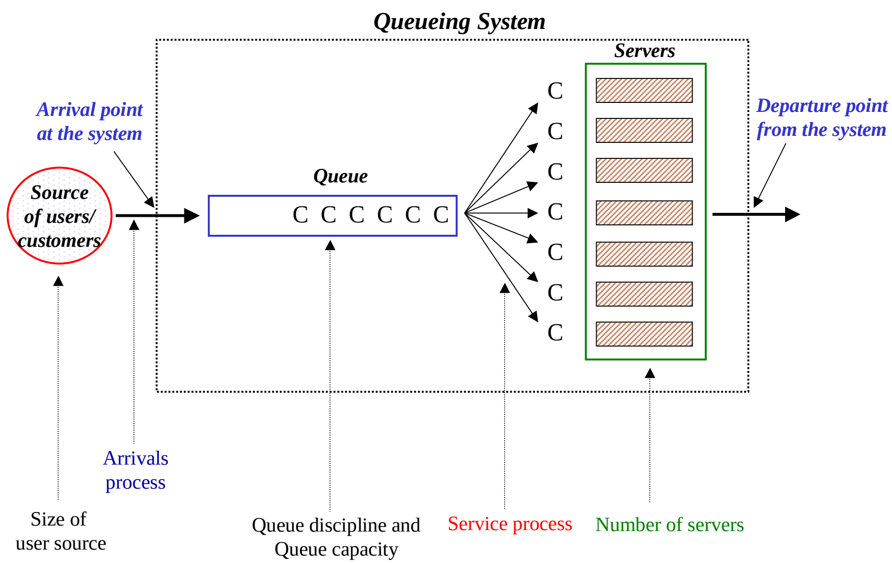
Queueing Models Can Be Essential in Analysis of Capital Investments



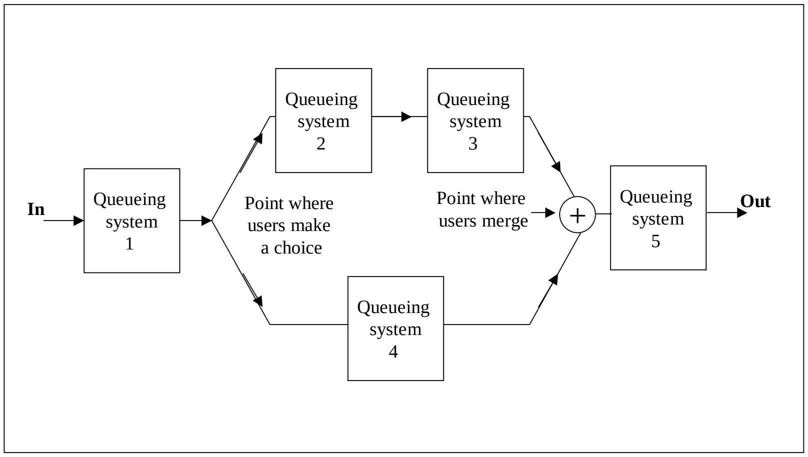
Strengths and Weaknesses of Queueing Theory

- Queueing models necessarily involve approximations and simplification of reality
- Results give a sense of order of magnitude, of changes relative to a baseline, of promising directions in which to move
- Closed-form results are essentially limited to “steady state” conditions and derived primarily (but not solely) for birth-and-death systems and “phase” systems
- Some useful bounds for more general systems at steady state
- Numerical solutions are increasingly viable for dynamic systems

Queueing Process and Queueing System

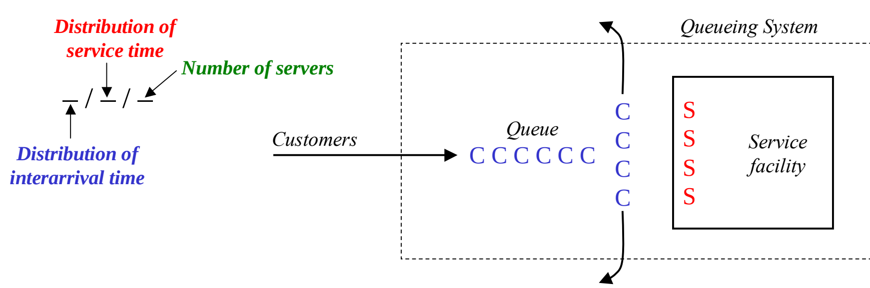


Queueing network consisting of five queueing systems



1.225, 11/26/02 — Lecture 8, Page 7

A Code for Queueing Models: *A/B/m*



- Some standard code letters for *A* and *B*:
  - *M: Negative exponential* (*M stands for memoryless*)
  - *D: Deterministic*
  - E_k:*kth-order Erlang distribution*
  - *G: General distribution*
- Model covered in this lecture: *M/M/1*

1.225, 11/26/02 — Lecture 8, Page 8

Terminology and Notation

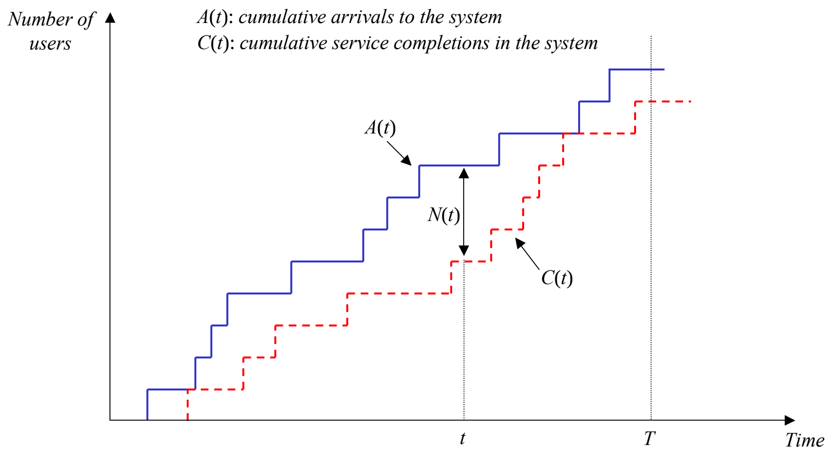
- *State of system*: number of customers in queueing system
- *Queue length*: number of customers waiting for service
- $N(t)$ = number of customers in queueing system at time t
- $P_n(t)$ = probability that $N(t)$ is equal to n
- λ_n: mean arrival rate of new customer when $N(t) = n$
- μ_n: mean (combined) service rate when $N(t) = n$
- *Transient condition*: state of system at t depends on the state of the system at t=0 or on t
- *Steady state condition*: system is independent of initial state and t
- s: number of servers (parallel service channels)
- If λ_n and the service rate per busy server are constant, then $\lambda_n=\lambda$, $\mu_n=s\mu$
- Expected interarrival time = $\frac{1}{\lambda}$
- Expected service time = $\frac{1}{\mu}$

Quantities of Interest at Steady State

- ***Given:***
  - λ = arrival rate
  - μ = service rate per service channel (number of servers =1, in this lecture)
- ***Unknowns:***
  - L = expected number of users in queueing system
  - L_q = expected number of users in queue
  - W = expected time in queueing system per user ($W = E(w)$)
  - W_q = expected waiting time in queue per user ($W_q = E(w_q)$)

- 4 unknowns $\Rightarrow$ We need 4 equations

Little's Law



$$L_T = \frac{\int_0^T N(t)dt}{T} = \frac{A(T)}{T} \cdot \frac{\int_0^T N(t)dt}{A(T)} = \lambda_T \cdot W_T$$

Relationships between L, L_q, W, and W_q

- 4 unknowns: L, W, L_q, W_q
- Need 4 equations. We have the following 3 equations:
  - $L = \lambda W$ (Little's law)
  - $L_q = \lambda W_q$
  - $W = W_q + \frac{1}{\mu}$
- If we know L (or any one of the four expected values), we can determine the value of the other three
- The determination of L may be hard or easy depending on the type of queueing model at hand (i.e. *M/M/1*, *M/M/s,* etc.)
- $L = \sum_{n=0}^{\infty} nP_n$ (P_n : probability that n customers are in the system)

Modeling Interarrival Time and Service Time

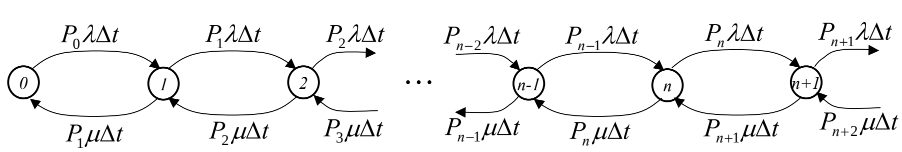
- T : Interarrival (service) time random variable
- Density function : $f_T(t) = \begin{cases} \alpha e^{-\alpha t} & ,t \geq 0 \\ 0 & ,t < 0 \end{cases}$
- $P\{0 \leq T \leq t\} = 1 - e^{-\alpha t}, \quad E(T) = \frac{1}{\alpha}, \quad \text{var}(T) = \frac{1}{\alpha^2}$
- For small Δt, $P\{0 \leq T \leq \Delta t\} \approx \alpha \Delta t$ (why?)
- $e^x = 1 + x + \sum_{k=2}^{\infty} \frac{x^k}{k!}$
- $P\{0 \leq T \leq \Delta t\} = 1 - e^{-\alpha \Delta t} = 1 - (1 - \alpha \Delta t + \sum_{k=2}^{\infty} \frac{(-\alpha \Delta t)^k}{k!})$

  $\approx \alpha \Delta t$ (for small Δt)
- Interarrival Time : $\alpha = \lambda$; Service Time : $\alpha = \mu$

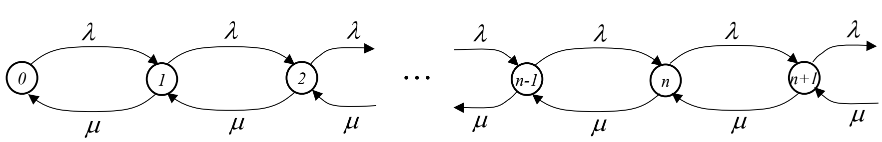
State Transition Diagram for *M/M/1*

- States:

- During Δt:



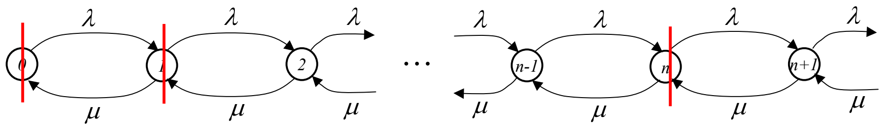
- Another way to represent it: *State Transition Diagram*



Observing State Transition Diagram from Two Points

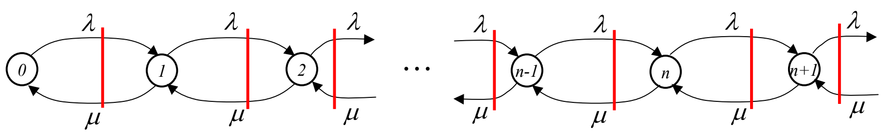
- From point 1:

$\lambda P_0 = \mu P_1 \quad (\lambda+\mu)P_1 = \lambda P_0 + \mu P_2 \quad (\lambda+\mu)P_n = \lambda P_{n-1} + \mu P_{n+1}$



- From point 2:

$\lambda P_0 = \mu P_1 \quad \lambda P_1 = \mu P_2 \quad \lambda P_n = \mu P_{n+1}$



Derivation of P_0 and P_n

- Putting it all together: $P_1 = \frac{\lambda}{\mu}P_0, \quad P_2 = \left(\frac{\lambda}{\mu}\right)^2 P_0, \cdots, \quad P_n = \left(\frac{\lambda}{\mu}\right)^n P_0$
- Since $\sum_{n=0}^{\infty} P_n = 1, \quad \Rightarrow \quad P_0 \sum_{n=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^n = 1 \quad \Rightarrow \quad P_0 = \frac{1}{\sum_{n=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^n}$
- Let $\rho = \frac{\lambda}{\mu}$, then $\sum_{n=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^n = \sum_{n=0}^{\infty}\rho^n = \frac{1-\rho^{\infty}}{1-\rho} = \frac{1}{1-\rho} \quad (\because \rho < 1)$
- Therefore, $P_0 = \frac{1}{\sum_{n=0}^{\infty}\rho^n} = 1-\rho$ and $P_n = \rho^n(1-\rho)$

Derivation of L, W, W_q, and L_q

- $L = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^{\infty} n\rho^n (1-\rho) = (1-\rho)\sum_{n=0}^{\infty} n\rho^n = (1-\rho)\rho\sum_{n=1}^{\infty} n\rho^{n-1}$

$$= (1-\rho)\rho\frac{d}{d\rho}\left(\sum_{n=0}^{\infty}\rho^n\right) = (1-\rho)\rho\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right)$$

$$= (1-\rho)\rho\left(\frac{1}{(1-\rho)^2}\right) = \frac{\rho}{(1-\rho)} = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\lambda}{\mu-\lambda}$$

- $W = \frac{L}{\lambda} = \frac{\lambda}{\mu-\lambda}\cdot\frac{1}{\lambda} = \frac{1}{\mu-\lambda}$

- $W_q = W - \frac{1}{\mu} = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)}$

- $L_q = \lambda W_q = \lambda\cdot\frac{\lambda}{\mu(\mu-\lambda)} = \frac{\lambda^2}{\mu(\mu-\lambda)}$

Lecture 8 Summary

- ❑ Introduction to Queueing
- ❑ Conceptual Representation of Queueing Systems
- ❑ Codes for Queueing Models
- ❑ Terminology and Notation
- ❑ Little’s Law and Basic Relationships
- ❑ Exponential Distribution for Interarrival and Service times Modeling
- ❑ State Transition Diagram
- ❑ Derivation of waiting characteristics for *M/M/1*